

How many Conflicts does it need to be Unsatisfiable?

Dominik Scheder and Philipp Zumstein *

Institute of Theoretical Computer Science, ETH Zürich
8092 Zürich, Switzerland
dscheder@inf.ethz.ch, zuphilip@inf.ethz.ch

Abstract. A pair of clauses in a CNF formula constitutes a conflict if there is a variable that occurs positively in one clause and negatively in the other. Clearly, a CNF formula has to have conflicts in order to be unsatisfiable—in fact, there have to be many conflicts, and it is the goal of this paper to quantify how many.

An unsatisfiable k -CNF has at least 2^k clauses; a lower bound of 2^k for the number of conflicts follows easily. We improve on this trivial bound by showing that an unsatisfiable k -CNF formula requires $\Omega(2.32^k)$ conflicts. On the other hand there exist unsatisfiable k -CNF formulas with $O(\frac{4^k \log^3 k}{k})$ conflicts. This improves the simple bound $O(4^k)$ arising from the unsatisfiable k -CNF formula with the minimum number of clauses.

Keywords: satisfiability, unsatisfiable formulas, conflict graph, Lovász Local Lemma

1 Introduction

If you want to explain to your non-computer science friend what satisfiability of CNF formulas is all about, you will probably say it is about a list of constraints, all of which you want to satisfy simultaneously. Perhaps you will add that while each constraint is very easy to satisfy individually, the difficulty arises because many constraints conflict with each other. A natural guess is that if you cannot satisfy all your constraints, then there must be a lot of conflicts between them.

In our case, constraints are boolean clauses with no repetition of literals and no complementary literals, e.g., $x \vee \bar{y} \vee z$. A k -CNF formula is the conjunction of such clauses each containing *exactly* k literals. We use this notation for CNF formulas to be closer to the semantical viewpoint of formulas instead of using the notation of clause-sets which is closer to the syntactical viewpoint.

Two clauses have a *conflict* if one contains a positive literal, while the other contains its negation. Kullmann [1] introduced the notion of the symmetric conflict matrix, which has an entry for each pair of clauses counting the number of conflicts between them. We take the 0-1-version of this matrix interpreted as a graph. More formally, the *conflict graph* $CG(F)$ of a k -CNF formula F contains

* Research is supported by the SNF Grant 200021-118001/1

II

the clauses as vertices, and two clauses are connected if there is a conflict between them. Any lower bound on the number of edges in the conflict graph, i.e., the number of conflicts, is also a lower bound on the number of multi-edges in the symmetric conflict matrix interpreted as a multi-graph.

We study the extremal values of several natural parameters of $CG(F)$ for unsatisfiable k -CNF formulas F , such as its minimum degree, maximum degree, and of course, above all, its number of edges. We will use the notations d_{\min} , d_{\max} , e , respectively, for these parameters and use concepts of graph theory also in the context of CNF formulas, e.g., the neighborhood of a clause C in a CNF formula F , denoted by $\Gamma_F(C)$, is the set of all clauses in F conflicting with C . To avoid notational confusion we use the e for the number of edges and e for the Eulerian constant.

Further, we introduce a notation that will come handy when defining CNF formulas: Let $F = C_1 \wedge \dots \wedge C_m$ and $G = D_1 \wedge \dots \wedge D_n$ be two CNF formulas over disjoint sets of variables. Define

$$F \vee\!\!\!\bigvee G := \bigwedge_{\substack{i=1,\dots,m \\ j=1,\dots,n}} C_i \vee D_j.$$

The formula $F \vee\!\!\!\bigvee G$ is a CNF formula with mn clauses. If F is a k -CNF formula and G an ℓ -CNF formula, $F \vee\!\!\!\bigvee G$ is a $(k + \ell)$ -CNF formula. Moreover, by using distributivity it is easy to see that the two formulas $F \vee G$ and $F \vee\!\!\!\bigvee G$ are equivalent (describing the same boolean function) but $F \vee G$ is syntactically not a CNF formula. Note that a CNF formula is just *one* representation of a boolean function and there might exist other representations (logically equivalent formulas). The problem of deciding whether a given CNF formula is unsatisfiable is the same as to answer the question whether it is logically equivalent to the constant 0 function. Thus it is essential how a formula is represented.

Example 1. The *complete formula* K_k on the variables x_1, \dots, x_k is the k -CNF formula with all 2^k possible k -clauses on these variables. More formally

$$K_1 := x_1 \wedge \bar{x}_1, \quad K_{k+1} := (x_{k+1} \vee\!\!\!\bigvee K_k) \wedge (\bar{x}_{k+1} \vee\!\!\!\bigvee K_k).$$

K_k is a k -CNF formula, and using induction and the fact that $\vee\!\!\!\bigvee$ is logically equivalent to \vee , one can easily see that it is unsatisfiable. Furthermore the conflict graph $CG(K_k)$ is a clique, has 2^k vertices and $\binom{2^k}{2}$ edges.

It is clear that an unsatisfiable 1-CNF formula contains two complimentary clauses and thereby at least one edge. After deleting all clauses of degree 0 in an unsatisfiable k -CNF formula, the so-obtained formula stays unsatisfiable. Therefore, we have at least 2^k clauses because this is the minimum number of clauses needed for a k -CNF formula to be unsatisfiable, and every clause has degree at least 1. Thus all unsatisfiable k -CNF formulas have $\Omega(2^k)$ many conflicts. Example 1 shows that there is an unsatisfiable k -CNF with $\Theta(4^k)$ many conflicts. What is the right order of magnitude for the number of conflicts needed for an unsatisfiable k -CNF formula?

There is a similar question in Ramsey theory: A graph G is H -Ramsey if every 2-coloring of its edges contains a monochromatic copy of H . The size Ramsey number $\hat{r}(H)$ asks for the minimum number of edges over all H -Ramsey graphs where the Ramsey number $r(H)$ is the minimum number of vertices over all H -Ramsey graphs. For which graphs does it hold that $\hat{r}(H) = \binom{r(H)}{2}$? It is known that equality holds for complete graphs [2] but clearly not for stars, cf. [3] for a survey about the size Ramsey number.

1.1 Results

The technical groundwork, namely the Lopsided Lovász Local Lemma is discussed in Section 2. We examine thereby also the maximum and minimum degree of an unsatisfiable k -CNF formula. The main theorem follows in Subsection 3.1 and states that every unsatisfiable k -CNF formula has at least $\Omega(2.32^k)$ conflicts. A construction by Hoory and Szeider [4] shows that there exists unsatisfiable k -CNF formulas with $O(\frac{4^k \log^3 k}{k})$ conflicts, discussed in Subsection 3.2. In the end we discuss the maximization versions of these parameters and formulate some open problems.

2 Maximum Degree and Minimum Degree

The Lovász Local Lemma can be used to show that k -CNF formulas where every clause depends only on a small subset of the other clauses are always satisfiable, compare for example [5] Section 2.2. As an implication k -CNF formulas where every variable occurs only a few times are always satisfiable, cf. Corollary 5 and [6, 7]. The Lovász Local Lemma can also be used to define a branching rule for a DPLL-algorithm on SAT as shown in [8]. Another applications is shown in [9] for the MAX-SAT problem.

Definition 2. Let A_1, \dots, A_m be events in some probability space. A graph $G = (V, E)$ with $V = \{1, \dots, m\}$ is called a lopsided dependency graph if for any A_i and any $U \subseteq V \setminus (\{A_i\} \cup \Gamma_G(A_i))$ with $\Pr[\bigcap_{j \in U} \bar{A}_j] > 0$, it holds that

$$\Pr \left[A_i \mid \bigcap_{j \in U} \bar{A}_j \right] \leq \Pr [A_i] .$$

Lemma 3 ([10],[11],[12]). Let A_1, \dots, A_m be events in some probability space, and let G be a lopsided dependency graph for them. If there are numbers $0 \leq \gamma_i < 1$, $1 \leq i \leq m$, such that for any i ,

$$\Pr [A_i] \leq \gamma_i \prod_{j \in \Gamma_G(A_i)} (1 - \gamma_j) ,$$

then

$$\Pr [\bar{A}_1 \cap \dots \cap \bar{A}_m] > 0 .$$

Now think of a CNF formula $F = C_1 \wedge \dots \wedge C_m$. Set each variable independently uniformly at random. Define A_i to be the event that C_i is *not* satisfied. It is not difficult to see that the conflict graph of F is a lopsided dependency graph for the events A_1, \dots, A_m . Applying Lemma 3, we obtain the following result:

Theorem 4. *The maximum degree of any unsatisfiable k -CNF formula is at least $\frac{2^k}{e}$.*

Proof. The proof is basically given in [6] by Kratochvíl et al. Assume that the maximum degree of a k -CNF formula F is at most $d := \frac{2^k}{e} - 1$. Set each variable independently uniformly at random. Then for each clause C_i , the probability that it is not satisfied is 2^{-k} . Now apply Lemma 3 with $\gamma_i := \frac{1}{d+1}$ for all i , and use the fact that $(1 - \frac{1}{d+1})^d \geq e^{-1}$. \square

The complete formula K_k (Example 1) has maximum degree $2^k - 1$ and shows that this bound is tight up to a constant factor.

Corollary 5 (Kratochvíl et al. [6]). *Suppose F is a k -CNF formula. If every variable occurs in at most $\frac{2^k}{ek}$ clauses, then F is satisfiable.*

Proof. Consider any clause C of F . Clearly, every literal in C causes at most $\frac{2^k}{ek} - 1$ conflicts and hence the maximum degree of F is at most $\frac{2^k}{e} - k$. Thus F is satisfiable by Theorem 4. \square

Does this result implies anything about the number of conflicts in an unsatisfiable k -CNF formula? The number of conflicts is trivially at least the maximum degree, hence $\Omega(2^k)$. Further, some variable x occurs in many clauses of F . Assume that this variable is more or less *balanced*, i.e., it occurs equally often as a positive and negative literal. In this case this variable by itself induces $\Omega\left(\frac{4^k}{k^2}\right)$ conflicts. Does every unsatisfiable k -CNF formula have such a balanced high-frequency variable? The next example gives the most negative answer to this question: We will define an unsatisfiable formula in which every variable occurs *exactly once* negative.

Example 6. We set $F_1 := x_1 \wedge \bar{x}_1$, and for $k \geq 1$ define recursively

$$F_k := (\bar{x}_1 \vee \bar{x}_2 \vee \dots \vee \bar{x}_k) \wedge \bigwedge_{i=1}^k \left(F_{k-1}^{(i)} \vee x_i \right),$$

where $F_{k-1}^{(i)}$ are copies of F_{k-1} on different set of variables. By definition of the operator \vee one sees inductively that F_k is indeed a k -CNF formula. It is easy to see by using induction again that F_k is unsatisfiable for all $k \geq 1$ and that every negative literal occurs only once.

Let $a(k)$ be the number of clauses in F_k . We have $a(1) = 2$ and $a(k) = ka(k-1) + 1$. Solving this recurrence, we obtain $a(k) = \sum_{j=0}^k \frac{k!}{j!} = \lfloor ek! \rfloor$. Each “top level” variable x_i occurs once negatively and $\lfloor e(k-1)! \rfloor$ times positively.

Therefore, the number of conflicts in this formula is huge and does not give us any good upper bound. It seems that we have to pay for extreme non-balancedness by huge variable frequency. Still, the concept of balanced and non-balanced variables will be of great importance in the next section, when we prove a lower bound on the number of conflicts.

Let us make a second stab on proving a lower bound on $e(F)$ for F being an unsatisfiable k -CNF formula. Observe that $e(F) \geq \frac{d_{\min}(F)}{2} m(F)$, where $m(F)$ is the number of clauses. We know that $m(F) \geq 2^k$. The minimum degree $d_{\min}(F)$ gets 0 when we add clauses of degree 0. Thus to get any meaningful bound, we have to consider minimal unsatisfiable formulas. A CNF formula is called *minimal unsatisfiable* if it is unsatisfiable and deleting any clause makes it satisfiable. Minimal unsatisfiable formulas are interesting objects themselves, for they have many algorithmic aspects, which were studied for example in [13–17]. In this paper however, we use only some straightforward combinatorial properties. First, we can assume w.l.o.g. that F is a minimal unsatisfiable k -CNF formula, as every unsatisfiable CNF formula has a minimal unsatisfiable subformula (which does not have more conflicts). Second, we actually can state a lower bound on $d_{\min}(F)$ if F is minimal unsatisfiable:

Lemma 7. *If F is a minimal unsatisfiable k -CNF formula, then $d_{\min}(F) \geq k$.*

Proof. Let F be a minimal unsatisfiable k -CNF formula and assume for contradiction that there is a clause C such that its degree in the conflict graph is less than k . By minimality of F , $F - C$ is satisfiable. Take a satisfying assignment α of $F - C$. The neighbors $\Gamma_F(C)$ of C are satisfied by α , so we can assign to each $D \in \Gamma_F(C)$ a variable x_D such that the corresponding literal in D is satisfied. At least one of the k variables in C is not assigned, say ℓ is the corresponding literal in C . By changing α such that ℓ is true, we get a satisfying assignment for F , which is a contradiction. \square

This result is also tight: The k -CNF formula F_k defined above is minimal unsatisfiable, and has a minimum degree of k .

Corollary 8. *Every unsatisfiable k -CNF formula has at least $k \cdot 2^{k-1}$ conflicts.*

3 Number of Conflicts

3.1 A Lower Bound

In this section, we prove a lower bound on the number of conflicts in an unsatisfiable k -CNF formula. Our main technical tool will be a corollary of the Lopsided Lovász Local Lemma:

Corollary 9. *Let F be a CNF formula not containing the empty clause. If for all clauses C in F it holds that*

$$\sum_{D \in \Gamma_F(C)} \Pr [D \text{ not satisfied}] \leq \frac{1}{4}, \quad (1)$$

then F is satisfiable.

Proof. Write $F = C_1 \wedge \dots \wedge C_m$. First, we can assume that there are no isolated clauses, i.e., every clause has some conflict. We apply Lemma 3 with $\gamma_i := 2 \Pr[C_i \text{ not satisfied}]$. Note that (1) implies $\Pr[C_i \text{ not satisfied}] \leq \frac{1}{4}$ for each C_i , thus $\gamma_i \leq \frac{1}{2}$ for all i . A short calculation using the fact that $\prod(1 - \gamma_j) \geq 1 - \sum \gamma_j$ completes the proof. \square

Consider the function $f(p) = 1 - \log(1 - p) - \frac{\log(1-p)}{\log(p)}$ which has a unique $p^* \in (0, \frac{1}{2})$ maximizing $f(p)$. In fact, $p^* \approx 0.30$ and $f(p^*) > 1.218$.

Theorem 10. *Let F be an unsatisfiable k -CNF formula. For any $p < \frac{1}{2}$, F has at least*

$$\frac{2^{f(p)k}}{8 + 128k^2}$$

conflicts. Furthermore, plugging in the optimal value p^ , we obtain that F has $\Omega(2.32^k)$ conflicts.*

Before we proceed to the proof, we explain the basic idea behind it. The Lovász Local Lemma implies that any unsatisfiable k -CNF formula contains a variable of high degree, namely around $\frac{2^k}{\theta k}$. Assume that this variable is more or less *balanced*, i.e., it occurs positively and negatively equally often. In this case there are already $\Omega\left(\frac{4^k}{k^2}\right)$ conflicts due to this variable, and we are done. Otherwise, we can assume that if a literal u occurs frequently, then \bar{u} does not. A natural idea now comes to mind: In the probability space used in the Lovász Local Lemma, we set each variable independently, choosing 1 with probability $\frac{1}{2}$. Should we not bias u towards 1, if many clauses benefit from this, and only few suffer?

Proof (of Theorem 10). For a literal u , let $\text{occ}_F(u)$ denote the number of clauses in F containing the literal u . Please note that $\text{occ}_F(u)$ and $\text{occ}_F(\bar{u})$ may differ. Choose parameters ℓ, θ as follows: $\ell := \frac{k \log(1-p)}{\log(p)}$ and $\theta = \frac{2^{k-\ell}}{8k}$.

We will color the literals of the formula and choose a probability with which they are set to 1. (i) If both $\text{occ}_F(u) \geq \theta$ and $\text{occ}_F(\bar{u}) \geq \theta$, color u and \bar{u} red, and set each to 1 with probability $\frac{1}{2}$. (ii) If $\text{occ}_F(u) \geq \theta$ and $\text{occ}_F(\bar{u}) < \theta$, color u green and \bar{u} red, and set u to 1 with probability $1 - p$ (and thus \bar{u} to 1 with probability p). (iii) If both $\text{occ}_F(u) < \theta$ and $\text{occ}_F(\bar{u}) < \theta$, color both *black*,¹ and set each to 1 with probability $\frac{1}{2}$. What you should keep in mind is that every black literal is “not frequent”, and the complement of each red literal is “frequent”, and that green literals are likely to be satisfied, since $1 - p > \frac{1}{2}$.

Let $e(F)$ be number of conflicts/edges in F . As a first observation, note that any literal u causes $\text{occ}_F(u)\text{occ}_F(\bar{u})$ conflicts. Summing this up over all red literals u we obtain the number of conflicts caused by red literals. However, we might

¹ or white, if you are working on a blackboard. . .

(i) count the pair $\{u, \bar{u}\}$ twice, if both are red, and (ii) count the same conflict k times, as two clauses can have up to k conflicting literals, e.g. $(u_1 \vee \dots \vee u_k)$ and $(\bar{u}_1 \vee \dots \vee \bar{u}_k)$. Still, we obtain

$$\begin{aligned} e(F) &\geq \frac{1}{2k} \sum_{u:\text{red}} \text{occ}_F(u) \text{occ}_F(\bar{u}) \geq \frac{\theta}{2k} \sum_{u:\text{red}} \text{occ}_F(u) \\ &\geq \frac{\theta}{2k} |\{C \in F \mid C \text{ contains at least one red literal}\}|. \end{aligned} \quad (2)$$

Our interpretation of this inequality is that if F has few conflicts, then only few clauses can contain red literals. Next, we define a new formula F' as follows: Start with F , and for each clause C in F that has fewer than ℓ green literals, and no red ones, remove all green literals. This formula F' is still unsatisfiable, and $e(F) \geq e(F')$. It is no k -CNF formula anymore, but each clause has at least $(k - \ell)$ literals. Further, $\text{occ}_{F'}(u) \leq \text{occ}_F(u)$ for any literal u .

Since F' is unsatisfiable, we can use the contrapositive of Corollary 9, i.e., there is a clause C^* in F' such that

$$\sum_{D \in \Gamma_{F'}(C^*)} \Pr[D \text{ not satisfied}] > \frac{1}{4}. \quad (3)$$

Let us partition $\Gamma_{F'}(C^*)$ into sets \mathcal{B} , \mathcal{BG} and \mathcal{BGR} as follows: Let \mathcal{B} contain all clauses $D \in \Gamma_{F'}(C^*)$ containing only black literals, \mathcal{BG} those containing at least one green literal, but no red ones, and \mathcal{BGR} those containing at least one red literal. There are several useful observations: First, by construction of F' , every clause in \mathcal{BG} contains at least ℓ green literals. Second, clauses in \mathcal{B} have at least $k - \ell$ literals, and \mathcal{BGR} and \mathcal{BG} contain only k -clauses. We can give certain bounds on $|\mathcal{BGR}|$ and $|\mathcal{B}|$. Every $D \in \mathcal{BGR}$ contains a red literal, hence by (2)

$$|\mathcal{BGR}| \leq \frac{2ke(F)}{\theta}. \quad (4)$$

To give a bound on $|\mathcal{B}|$, note that for every $D \in \mathcal{B}$ there is a black literal u in C^* such that \bar{u} is in D . Since $\text{occ}_{F'}(\bar{u}) \leq \theta$ for each such literal u in C^* , and C^* contains at most k of them,

$$|\mathcal{B}| \leq k\theta. \quad (5)$$

To evaluate the sum in (3), let us estimate the probabilities with which the clauses in the three sets are unsatisfied. First, a clause $D \in \mathcal{BGR}$ has at least k literals, and each is satisfied with probability at least p (we pessimistically assume the worst case, namely that all literals in D are red). We obtain

$$\forall D \in \mathcal{BGR} : \Pr[D \text{ not satisfied}] \leq (1 - p)^k. \quad (6)$$

For $D \in \mathcal{B}$, observe that D has at least $k - \ell$ literals, which are all black, thus

$$\forall D \in \mathcal{B} : \Pr[D \text{ not satisfied}] \leq 2^{-k+\ell}. \quad (7)$$

VIII

For $D \in \mathcal{BG}$, we know that D contains at least ℓ green literals, each of which is satisfied with probability $1 - p$. All other literals in D are black. Therefore we obtain

$$\forall D \in \mathcal{BG} : \Pr[D \text{ not satisfied}] \leq 2^{-k+\ell} p^\ell . \quad (8)$$

We plug (4)–(8) into (3) and get

$$\frac{1}{4} < |\mathcal{BG}| 2^{-k+\ell} p^\ell + 2ke(F)\theta^{-1}(1-p)^k + k\theta 2^{-k+\ell} .$$

This yields a lower bound on $|\mathcal{BG}|$. Since every clause in \mathcal{BG} has a conflict with C , it yields a lower bound on $e(F')$, and thus on $e(F)$:

$$e(F) \geq |\mathcal{BG}| > p^{-\ell} (2^{k-\ell-2} - 2ke(F)\theta^{-1}(1-p)^k 2^{k-\ell} - k\theta) .$$

Plugging in our values for θ and ℓ , a few calculations show that

$$e(F) > \frac{2^{f(p)k}}{8 + 128k^2} .$$

This completes the proof of the theorem. \square

We should make some comments on the proof above. It is a natural idea to choose a probability $\neq \frac{1}{2}$ for unbalanced variables. However, it is not clear why it makes sense to delete green (frequent) literals from clauses. Are these not exactly those literals making the clause more likely to be satisfied? What would happen if we re-did the proof without deleting green literals? We would have to assume the worst case that every $D \in \mathcal{BG}$ contains $k - 1$ black literals and only one green, only marginally pushing up its probability of being satisfied. Overall, we would obtain a bound like $\Omega\left(\frac{k^2 2^k}{\ln k}\right)$, if we choose our parameters optimally. The benefit of crossing out some green literal becomes clear if one thinks the other way round: We are given a formula with fewer than $\frac{2^{f(p)k}}{8+128k^2}$ conflicts and want to prove that it is satisfiable. By deleting green literals from some clauses D , we sacrifice by decreasing the probability of D being satisfied. On the other hand, no clause C can have many of these shrunk, now completely black clauses D in its neighborhood, due to the bound $|\mathcal{B}|$. We are actually making the conflict graph much sparser, thus more amenable to the Local Lemma.

3.2 Upper Bound

For an upper bound construction we show that the construction given by Hoory and Szeider [4] gives an unsatisfiable k -CNF formula with $O\left(\frac{4^k \log^3 k}{k}\right)$ conflicts. They used this construction to give an instance of a k -CNF formula with few occurrences of each variable.

For notational matter we write $K(v_1, \dots, v_s)$ to denote the complete formula on the variables v_1, \dots, v_s , and $K^-(v_1, \dots, v_s)$ is used for the formula

obtained by $K(v_1, \dots, v_s)$ deleting the all-positive-clause $(v_1 \vee \dots \vee v_s)$. The formula $K^-(v_1, \dots, v_s)$ is equivalent to $(\bar{v}_1 \wedge \dots \wedge \bar{v}_s)$. Let $k \geq 1$ and choose $\ell = \lfloor \log(k/\log^2 k) \rfloor$, $u = \lfloor k/\ell \rfloor$, and $v = k - lu$. Define the formula $F = \bigwedge_{i=0}^u F_i$ where

$$F_0 = K(z_1, \dots, z_v) \vee \bigvee_{i=1}^u K^-(x_1^{(i)}, \dots, x_1^{(i)}),$$

$$F_i = K(y_1^{(i)}, \dots, y_{k-\ell}^{(i)}) \vee (x_1^{(i)} \vee \dots \vee x_\ell^{(i)}).$$

Replace the complete formulas by the constant 0 function and the ‘‘almost complete’’ formulas $K^-(x_1^{(i)}, \dots, x_s^{(i)})$ by $(\bar{x}_1^{(i)} \wedge \dots \wedge \bar{x}_s^{(i)})$ to see that this formula is unsatisfiable. It remains to calculate the number of conflicts in this formula:

$$\begin{aligned} e(F) &= e(F_0) + ue(F_1) + ue(F_0, F_1) \leq |F_0|^2 + u|F_1|^2 + u|F_0||F_1| \\ &\leq \frac{e^2 4^k}{k^2} + \frac{4^k \log^3 k}{k \log^2 e} + \frac{4^k e \log^3 k}{k \log^2 e} = O\left(\frac{4^k \log^3 k}{k}\right). \end{aligned}$$

The calculations for $|F_0|, |F_1|$ are made in [4] and the better upper bound described there can also be used to get rid of one logarithm, i.e., this would yield an unsatisfiable k -CNF formula with $O\left(\frac{4^k \log^2 k}{k}\right)$ conflicts. The details are omitted because the lower bound is far away from this upper bound and we cannot be sure what the right order of magnitude is.

4 Maximizing the Parameters

How big can the parameters d_{\min}, d_{\max}, e of a minimal unsatisfiable k -CNF formula be? We will show that there exists minimal unsatisfiable k -CNF formulas with arbitrarily large minimum degree, for $k \geq 3$. From this it follows that the maximum degree, number of edges are arbitrarily big as well.

Example 11. The *cycle formula* C_ℓ on the variables x_1, \dots, x_ℓ contains the clauses $(\bar{x}_i \vee x_{i+1}), i = 1, \dots, \ell$ (the index is taken here and in the following modulo ℓ). We can interpret each clause as a logical implication $x_i \rightarrow x_{i+1}$ or as an inequality $x_i \leq x_{i+1}$ over the boolean values 0, 1. It follows that C_ℓ is satisfied if and only if all the variables x_1, \dots, x_ℓ are set to the same value. By adding the clauses $(x_1 \vee x_2), (\bar{x}_1 \vee \bar{x}_2)$ to C_ℓ we obtain a minimal unsatisfiable 2-CNF formula. Consider now

$$H_{k,\ell} := (C_\ell \wedge (x_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_2)) \vee K(y_1, \dots, y_{k-2})$$

It is not difficult to see for two minimal unsatisfiable formulas F, G also $F \vee G$ is minimal unsatisfiable. Thus $H_{k,\ell}$ is a minimal unsatisfiable k -CNF formula. It contains $2^{k-2}(\ell + 2)$ clauses, and each clause has at least $(2^{k-2} - 1)(\ell + 2) + 2$ conflicts. Therefore, the minimum degree of a minimal unsatisfiable k -CNF formula can be arbitrarily large, for each fixed $k \geq 3$.

5 Conclusion

Let e_k denote the minimum number of conflicts/edges over all unsatisfiable k -CNF formulas. Theorem 10 and the construction in Section 3.2 leads to the following asymptotic bounds

$$\Omega(2.32^k) \leq e_k \leq O\left(\frac{4^k \log^3 k}{k}\right).$$

This improves over the trivial bounds, but the gap between the lower and upper bound is still huge. We suspect that the true value of e_k is much closer to the upper bound, but it took us a considerable effort to get away from lower bounds like $\Omega(k^t 2^k)$ for some fixed t . The right magnitude of e_k is therefore still the main open question of this paper.

It is easy to change the definition of the complete formula in such a way that every pair of clauses has exactly one conflicting variable. By substituting this variant of the complete formula into the example given in Section 3.2 yields an unsatisfiable k -CNF formula with $\Theta\left(\frac{4^k \log^3 k}{k}\right)$ multi-edges.

As we pointed out in Example 6 there exists unsatisfiable k -CNF formulas where all variables of high degree are very unbalanced. But even there it holds that there is one variable u such that it induces almost all conflicts, i.e., $\text{occ}(u) \cdot \text{occ}(\bar{u})$ is already very large. Is it true in general that the maximum of $\text{occ}(u) \cdot \text{occ}(\bar{u})$ is very large?

Acknowledgment. We thank all the participants of Gremo Workshop on Open Problems, GWOP 2007, for the helpful discussions at the workshop.

References

1. Kullmann, O.: The combinatorics of conflicts between clauses. In Giunchiglia, E., Tacchella, A., eds.: Theory and Applications of Satisfiability Testing, 6th International Conference, SAT 2003. Santa Margherita Ligure, Italy, May 5-8, 2003 Selected Revised Papers. Volume 2919 of Lecture Notes in Computer Science., Springer (2004) 426–440
2. Erdős, P., Faudree, R.J., Rousseau, C.C., Schelp, R.H.: The size ramsey number. *Periodica Mathematica Hungarica* **9**(2-2) (1978) 145–161
3. Faudree, R.J., Schelp, R.H.: A survey of results on the size Ramsey number. In: Paul Erdős and his mathematics, II (Budapest, 1999). Volume 11 of Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest (2002) 291–309
4. Hoory, S., Szeider, S.: A note on unsatisfiable k -CNF formulas with few occurrences per variable. *SIAM Journal on Discrete Mathematics* **20**(2) (2006) 523–528
5. Welzl, E.: Boolean satisfiability – combinatorics and algorithms (lecture notes) (2005) <http://www.inf.ethz.ch/~emo/SmallPieces/SAT.ps>.
6. Kratochvíl, J., Savický, P., Tuza, Z.: One more occurrence of variables makes satisfiability jump from trivial to NP-complete. *SIAM Journal of Computing* **22**(1) (1993) 203–210

7. Berman, P., Karpinski, M., Scott, A.D.: Approximation hardness and satisfiability of bounded occurrence instances of SAT. *Electronic Colloquium on Computational Complexity (ECCC)* **10**(022) (2003)
8. Hooker, J.N., Vinay, V.: Branching rules for satisfiability. *Journal of Automated Reasoning* **15** (1995) 359–383
9. Czumaj, A., Scheideler, C.: A new algorithm approach to the general Lovász local lemma with applications to scheduling and satisfiability problems (extended abstract). In: *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, New York, ACM (2000) 38–47 (electronic)
10. Erdős, P., Spencer, J.: Lopsided Lovász local lemma and Latin transversals. *Discrete Appl. Math.* **30**(2-3) (1991) 151–154 *ARIDAM III* (New Brunswick, NJ, 1988).
11. Alon, N., Spencer, J.H.: *The probabilistic method*. Second edn. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley-Interscience [John Wiley & Sons], New York (2000) With an appendix on the life and work of Paul Erdős.
12. Lu, L., Székely, L.: Using Lovász local lemma in the space of random injections. *Electron. J. Combin.* **14**(1) (2007) Research Paper 63, 13 pp. (electronic)
13. Papadimitriou, C.H., Wolfe, D.: The complexity of facets resolved. *J. Comput. Syst. Sci.* **37**(1) (1988) 2–13
14. Szeider, S.: Minimal unsatisfiable formulas with bounded clause-variable difference are fixed-parameter tractable. *J. Comput. Syst. Sci.* **69**(4) (2004) 656–674
15. Kleine Büning, H.: On subclasses of minimal unsatisfiable formulas. *Discrete Appl. Math.* **107**(1-3) (2000) 83–98 *Boolean functions and related problems*.
16. Fleischner, H., Kullmann, O., Szeider, S.: Polynomial-time recognition of minimal unsatisfiable formulas with fixed clause-variable difference. *Theoret. Comput. Sci.* **289**(1) (2002) 503–516
17. Kleine Büning, H., Zhao, X.: The complexity of some subclasses of minimal unsatisfiable formulas. *J. Satisf. Boolean Model. Comput.* **3**(1-2) (2007) 1–17